Compte rendu de la rencontre du C4 (Comité des Chercheurs Calculant au CINES) avec les représentants du CINES et de GENCI
Vendredi 13 mai 2016 - Montpellier

La réunion s'est déroulée en 2 parties, une première réunion le matin entre le C4 et le CINES, une deuxième l'après-midi en présence de GENCI. Pour rappel, le CINES (Centre Informatique National de l'Enseignement Supérieur) héberge actuellement la machine de calcul OCCIGEN.

1. Rencontre C4-CINES

Membres du C4 présents :

- CT5 : Virginie Grandgirard (présidente)
- CT1 : Sébastien Theetten
- CT2a/CT2b : Florent Duchaine
- CT7 : Jérôme Hénin
- CT8 : Sébastien Le Roux
- CT9 : Évelyne Lampin

Membres du CINES présents :

- Johanne Charpentier (Administrateur système OCCIGEN), Jean-Christophe Penalva (Support utilisateurs OCCIGEN), Mathieu Cloirec (Coordinateur formations), Philippe Prat (Chef de projet eDARI)
- Olivier Rouchon (Responsable Département Calcul Intensif)
- Francis Daumas (Directeur du CINES)

1.1 Point CINES :

- Une maintenance d'OCCIGEN est prévue le 7 juin.

- **Extension d'OCCIGEN** :
Il est prévu une extension « OCCIGEN 2 » de la machine OCCIGEN, avec pour objectif un quasi doublement de sa puissance. Pour mémoire, la machine dispose de 2106 nœuds soit 50 544 cœurs. Actuellement on ne connaît pas encore la configuration exacte, car les négociations pour l'achat sont toujours en cours. Les processeurs devraient avoir la même fréquence d'horloge, avec plus de cœurs par nœud. La machine sera mise en production via l'appel DARI à partir de la 1ère session 2017. Il y aura des interruptions à prévoir en fin d'année pour connecter l'extension à la machine actuelle. Des grands défis, c'est-à-dire des gros projets pour tester la machine avant sa mise en production, seront probablement organisés en novembre/décembre, selon des modalités de sélection encore à définir. Pour l'extension de la machine, refroidie par eau tiède comme OCCIGEN, le CINES réalise une augmentation de puissance électrique de 1MW dans la salle machine. Des améliorations vont également être apportées en ce qui concerne le refroidissement, via le confinement des allées froides/chaudes pour diminuer les pertes et améliorer le PUE. Rappelons que le refroidissement par eau tiède de la machine est possible car il s'effectue au plus près des microprocesseurs, et non par exemple sur les façades des armoires, ce qui devait se faire par eau froide. Cela permet donc de diminuer le coût, puisque l'eau doit être moins refroidie. Quand la machine sera en production, la soumission se fera sur OCCIGEN (Haswell) et OCCIGEN2 (Broadwell), comme elle se faisait sur JADE1 (Harpertown) et JADE2 (Nehalem) autrefois, mais avec moins de différentiel entre les deux tranches d'OCCIGEN qu'il n'y en avait entre celles de JADE.

- La date de l'inauguration d'OCCIGEN n'a toujours pas été fixée et la publication des grands défis d'OCCIGEN dans la gazette du CINES est donc bloquée.

- Dans le cadre de la cellule de veille technologique animée par GENCI il a été décidé d'installer au CINES un prototype basé sur des nouveaux nœuds à processeur XEON PHI (Knights Landing). Des tests internes auront lieu dans un 1er temps (sur une machine de type Bull Sequana), puis des contacts seront pris avec des utilisateurs identifiés avant d'ouvrir à la communauté. Il a été choisi de décorréler cet aspect de l'achat d'OCCIGEN2, car il semblait important d'effectuer un travail préliminaire avant d'intégrer ce type de processeurs aux moyens de calcul généralistes.

- En 2015, la charge de la machine OCCIGEN a été de 80 à 90 % tous les mois. La machine marche bien. Il n'y a pas énormément d'attente dans les queues effectivement, comme le confirme le C4. Peut-être est-ce dû au fait que les moyens de calcul ont nettement augmenté avec OCCIGEN, alors que le nombre de compte, et donc d'utilisateurs, est stable. Les types de jobs étaient initialement distribués selon une gaussienne centrée aux alentours de 500 cœurs.

Récemment, la distribution a évolué vers une log normale avec un pic à nombre de cœurs plus élevés. Ces gros jobs ne consomment pas tous forcément énormément, et pour certains peuvent être attribués à des tests de scalabilité, signe que les utilisateurs optimisent le nombre de cœurs utilisés avant de lancer une campagne, ce qui est salué par le CINES. Il y a une nette diminution des temps d'attente la nuit et les weekends, ainsi qu'en août, avis aux amateurs...

- La pression en demande (heures demandées sur heures disponibles) est passée de 150 % à 140 % en session 2.

1.2 Réponses aux questions utilisateurs relayées par les membres du C4

- Des utilisateurs ont remarqué une fluctuation des temps de restitutions jusqu'à un facteur 2. Le CINES travaille sur ce problème très délicat concernant le réseau d'interconnexion InfiniBand. Un outil (UFM) est en cours de déploiement pour obtenir des informations plus précises sur le fonctionnement et la performance des équipements Mellanox, afin de tester et continuer d'investiguer ce réseau InfiniBand.

- Il est maintenant possible sur quelques nœuds de lancer des jobs en mode partagé, ce qui signifie que plusieurs jobs sont effectués en même temps sur un même nœud, ce qui s'adresse à des jobs qui sont faiblement parallélisables, typiquement GAUSSIAN, qui n'est pas scalable au-delà de 1 nœud. Ce nouveau mode de fonctionnement permet une meilleure corrélation entre les heures consommées et les heures comptabilisées dans le DARI. Auparavant, pour un cœur réservé dans un nœud, tous les cœurs du nœud étaient comptabilisés. Avec cette nouvelle mesure, la consommation est enregistrée au prorata de la demande mémoire, et selon le nombre de cœurs assignés. Ce nouveau mécanisme permet aussi de libérer davantage de nœuds, puisque les petits jobs, ne « surconsommeront » plus les ressources.

- Le C4 interpelle le CINES sur la formation GAUSSIAN qu'il a organisée, et qui a déçu les participants. Il semble qu'à l'origine il y ait une incompréhension sur le but de la formation, qui a été pensée côté CINES pour des débutants, alors que l'attente des participants était plus poussée. Le CINES est prêt à organiser ou à participer à une nouvelle formation mieux adaptée à l'auditoire pour répondre à la demande, avec éventuellement une implication de GENCI, en commun avec les autres centres de calcul. Le CINES a installé le binaire du code fourni par GAUSSIAN, et travaille actuellement sur une compilation avec les compilateurs les mieux adaptés sur OCCIGEN (Intel). Le CINES sollicite les utilisateurs pour faire remonter leur satisfaction/insatisfaction de la version actuellement installée, et participer à l'évaluation et la validation du code compilé en interne.

- Le CCFR : Suite aux mécontentements de plusieurs utilisateurs concernant cette liaison 10 Gb entre centres nationaux le C4 avait demandé au CINES un point spécifique sur le sujet. Pour rappel l'ouverture de cette ligne était initialement prévue pour le 2e trimestre 2015. Ceci n'est toujours pas le cas car il persiste malheureusement des problèmes de performances entre le CINES et le CCRT. Le débit constaté à l'heure actuelle est plutôt de l'ordre de 1Gb. Il n'est donc pas possible d'ouvrir la ligne tant que les performances ne sont pas celles attendues entre tous les centres de calcul. Tous les tests possibles en interne au CINES ont été faits. Il semblerait que le problème soit extérieur. Des pistes sont à l'étude, comme un problème au niveau du filtrage du firewall Checkpoint (le CINES est en contact avec le constructeur de l'équipement sur ce point) et/ou un problème lié à l'équipement RENATER. Le C4 insiste sur le fait que ceci pénalise grandement certains utilisateurs. Le CINES est tout à fait conscient du problème et fait son maximum pour le résoudre dans les meilleurs délais.

- **Problème de quota mémoire :** Le C4 transmet la question d'un utilisateur sur un problème de taille de SCRATCH. Le quota par défaut est de 4 To, mais chaque utilisateur peut demander d'augmenter ce quota. Attention cet espace est temporaire et fragile, donc il sera d'abord demandé à l'utilisateur de nettoyer au maximum son espace SCRATCH, avant dans un deuxième temps de procéder à son extension, sachant que la capacité totale du SCRATCH est de 5 Po. Sur le STORE, le quota est volontairement très bas mais peut-être augmenter sans souci à la demande des utilisateurs, afin de s'adapter au mieux aux besoins de chacun. Par contre il y a une limite sur le nombre de fichiers. A terme un quota sera mis en place sur SCRATCH via des critères sur les dates d'utilisation des fichiers par exemple.

- **Echange de données entre différents projets :** Il est maintenant possible d'échanger des données entre 2 projets, soit via un échange ponctuel de fichiers réalisé par un agent du CINES, soit via des répertoires partagés entre les projets. Pour cela, il faut contacter le CINES (svp@cines.fr), en mettant les 2 porteurs de projets en copie. Il y aura un formulaire disponible pour cela d'ici peu.

- **Logins :** Une harmonisation tout au moins partielle entre les centres de calcul est en cours de discussion en ce qui concerne le renouvellement des logins. Il n'est malheureusement pas possible d'homogénéiser complètement, car les critères sont différents suivant les tutelles des centres de calcul et leurs procédures de sécurité. La piste envisagée au CINES serait un formulaire en ligne type CCRT, avec mémoire des informations d'une demande à l'autre. Le CINES insiste sur le fait que grâce à des négociations au niveau du SGDSN (Secrétariat Général de la Défense et de la Sécurité Nationale), on est passé d'un mécanisme beaucoup plus lourd, où toutes les demandes de logins devaient remonter au HFSD (Haut Fonctionnaire Sécurité Défense) avec un délai de traitement qui pouvait atteindre 2 mois, à la formule

actuelle d'examen au niveau des FSD des centres, en arguant que les utilisateurs des centres nationaux sont considérés comme des visiteurs.

- L'utilisation de clés d'identification n'est plus compatible avec la politique de sécurité des systèmes d'information. Cette interdiction fait suite à l'audit réalisé par l'ANSSI (Agence Nationale de la Sécurité des Systèmes d'Information) au CINES. Le CINES est autant gêné que les utilisateurs par cette restriction mais ne peut pas aller à l'encontre des préconisations de cet organisme.

- Suite à une demande des utilisateurs lors de la dernière réunion C4, le CINES travaille sur une commande ligne « etat_projet » (dans l'esprit du ccc_myproject sur Curie) qui permettra d'avoir directement accès sous UNIX à l'état de son projet (nombre d'heures déjà consommées, restantes, etc …). Cette commande devrait être disponible d'ici juin

- Une formation « outils de profiling » dans le cadre du PATC (PRACE) a été proposée et annulée car il n'y avait pas assez d'inscrits (outils Scorpe, scalasca). Le CINES se demande si le besoin existe encore dans les communautés (quatre formations de ce type ont été organisées en France dans un passé récent) et si les utilisateurs sont vraiment intéressés par ce genre de formation.

2. Rencontre C4-GENCI-CINES

Représentant GENCI:

- Arnaud Valois

- GENCI (Grand Equipement National de Calcul Intensif) va voir son périmètre revisité dans les mois qui viennent, pour inclure un volet « données » à ses activités.
Pour rappel, GENCI gère actuellement, sur le plan national, l'utilisation des machines de calcul qu'il a financé et confié aux centres de calcul via l'appel à projets DARI. L'évaluation des projets est effectuée par les experts des Comités Thématiques et l'attribution des heures aux utilisateurs par GENCI en consultation avec les directeurs des centres. La PDG de GENCI, Catherine Rivière, arrive en fin de mandat mi-juin. Elle n'a pas souhaitée le renouveler. Un nouveau PDG sera donc nommé très prochainement.
Certains présidents de CT (CT2b, 3, 4, 5, 6 et 10) arrivaient également en fin de mandat. Pour rappel, les mandats sont de 3 ans, renouvelables une seule fois. La liste des nouveaux présidents est la suivante :
  - CT2b (Écoulements réactifs et multiphasiques) : Nasser Darabiha - *Laboratoire EM2C - ECP & UPR CNRS*
  - CT3 (Biologie et santé) : Laurent Desbat - *Laboratoire TIMC-IMAG - UMR CNRS, Université Joseph Fourier*
  - CT4 (Astrophysique et géophysique) : Frédéric Bournaud - *Laboratoire AIM UMR CEA, CNRS*
  - CT5 (Physique théorique et physique des plasmas) : Eric Serre - *Laboratoire MP2P UMR CNRS, Université d'Aix-Marseille, École Centrale Marseille*
  - CT6 (Informatique, algorithmique et mathématiques) : Didier Auroux - *Laboratoire J.A. Dieudonné UMR CNRS, Université de Nice – Sophia Antipolis*
  - CT10 (Nouvelles applications et applications transverses du calcul intensif) : Bruno Scheurer – *CEA/DIF*

Le prochain renouvellement sera pour le CT9.
- Le C4 évoque le problème de non-uniformisation entre CT. Par exemple, certains experts demandent de regrouper les projets au niveau des groupes voire des laboratoires, alors que d'autres non. GENCI répond qu'il est laissé libre champ aux CT pour gérer cet aspect en fonction des thématiques.

- **Bilan 2015 DARI**: 1,1 milliard d'heures de calcul étaient disponibles, 1,5 milliard d'heures ont été demandées, 1.2 attribuées et 1 milliard consommées. Il y a une répartition homogène des demandes 2016 entre les 3 centres de calculs. La pression la moins forte est sur la machine Turing. Lors de la 1ère session spéciale Curie, un quart des demandes ont été sélectionnées en raison de la forte pression de la demande.

- **2e session de la campagne DARI 2016** : Le C4 fait remonter un manque de communication sur la 2ème session spéciale Curie, pour des calculs à faire entre septembre et décembre : l'appel n'a pas été fait isolément, mais en même temps que la 2e session DARI, ce qui fait que certains n'ont pas réalisé qu'il avait lieu. GENCI précise néanmoins qu'il avait été indiqué dans la note de cadrage, les mails d'ouverture de la campagne DARI (diffusés aux responsables de projets DARI et à la liste calcul) et la page d'accueil du site https://www.edari.fr, que les heures disponibles pour la session spéciale, issues de PRACE, avaient été ajoutées aux disponibilités de Curie pour la 2e session du DARI 2016, tout en faisant l'objet de conditions d'attribution spécifiques. GENCI insiste sur le fait qu'il faut bien consulter la note de cadrage des appels DARI, et justifie ce regroupement par des décisions tardives prises de part et d'autre, notamment en raison de contraintes externes liées à PRACE, ce qui n'a pas permis de communiquer l'information plus tôt à l'ensemble des utilisateurs.

- **Nouvelle organisation DARI pour 2017 :** GENCI présente la nouvelle organisation des appels DARI, car les sessions vont être décalées, et les durées vont changer, avec une période transitoire début 2017. Ce changement est fait

afin d'optimiser l'utilisation des supercalculateurs des centres de calcul, car il y a typiquement une sous-exploitation au mois de janvier, avec le retour des congés de fin d'année qui coïncide avec le démarrage des nouveaux projets et les demandes de renouvellement de login, ainsi qu'un déséquilibre observé dans l'utilisation de certains calculateurs entre le 1er et le 2nd semestre. Les sessions seront désormais avancées de 2 mois, pour démarrer le 1er mai et le 1er novembre, et les projets auront une durée de 1 an pour chaque session. Il sera toutefois toujours possible de faire des demandes complémentaires tous les 6 mois. Les appels à projets auront lieu en juillet/aout et février à partir de 2017 (l'appel en automne 2016 est maintenu pour un démarrage des allocations en janvier 2017). Les modalités de la période de transition pour atteindre ce régime ne sont pas encore fixées et elles font actuellement l'objet de discussions avec les représentants des utilisateurs des 3 centres de calcul nationaux et les présidents de comités thématiques. L'organisation de la période de transition devrait être figée au mois de juin et fera l'objet d'une communication la plus large possible afin de préparer au mieux cette période de transition en vue de l'appel à projets de l'automne 2016. GENCI et le C4 invite l'ensemble des utilisateurs à rester **vigilants sur les informations DARI**.

- GENCI travaille à assouplir les **modalités d'ouverture de compte,** avec l'objectif, à terme, de les uniformiser entre les 3 centres nationaux.

- **Rencontre C4-GENCI-CINES:** GENCI propose d'augmenter la fréquence des rencontres avec le C4 et le Cines, afin de pouvoir communiquer les informations importantes à un rythme équivalent à celui établi dans les deux autres centres de calcul (4 fois par an). La réunion du C4 étant complexe à organiser et mobilisant une journée entière, il estproposé d'organiser une réunion intermédiaire entre deux C4, au format allégé (max. 2h) et organisée par visio, de manière à augmenter la fréquence des échanges.

Prochaine réunion C4 prévue à l'automne. À cette occasion **les membres du C4 devront être renouvelés**. Un appel à candidature sera lancé d'ici Septembre 2016.